珍稀濒危飘带兜兰叶绿体全基因组种内变异研究

高鑫祯[1,2]，唐露[1]，汪雨[1,3]，邵士成[1]，罗艳[1*]

(1. 中国科学院西双版纳热带植物园，云南 勐腊 666303；2. 中国科学院大学，北京 100049；
3. 西南林业大学 园林园艺学院，昆明 650224)

摘　要：飘带兜兰 (*Paphiopedilum parishii*) 分布范围狭窄，仅在中国、缅甸、泰国以及老挝有少量分布。近年由于生境破坏和人为滥采导致飘带兜兰野生种群极度缩减。遗传多样性代表着物种适应环境和生存发展的能力，保存珍稀濒危物种的遗传多样性是物种保护的重要目标，但飘带兜兰的遗传信息还不清楚。该研究对飘带兜兰 4 个野生个体经测序、组装、注释获得的叶绿体基因组序列，与已公开的另外 2 个个体的叶绿体全基因组序列进行比较，分析种内差异。结果表明：(1) 飘带兜兰叶绿体基因组具有典型被子植物叶绿体基因组环状四分体结构，基因组长度为 154 403~154 809 bp，共编码 129 个基因，包括 78 个蛋白质编码基因、39 个 tRNA 基因、8 个 rRNA 基因，以及 4 个假基因。(2) 在飘带兜兰 6 个个体叶绿体基因组中检测到 103~107 个 SSRs (simple sequence repeats) 位点，其中 21 个 SSR 位点具有多态性。此外，6 个个体叶绿体基因组中还检测到 60 个长序列重复，包括 17~21 个正向重复，18~29 个反向重复，9~16 个回文重复，4~9 个互补重复。(3) 通过比较 6 个个体叶绿体基因组序列的核苷酸多样性，共发现 70 处变异，包括 10 个 SNPs (single nucleotide polymorphism) 以及 60 个插入缺失 (InDels)。其中有 3 个 SNP 位点发生了非同义替换，导致编码功能基因的氨基酸发生改变；19 个插入缺失多态性较高，具有开发为分子标记的潜力。(4) 通过计算核苷酸多样性值 (P_i) 共发现 8 个有变异的区域，P_i 值为 0 ~ 0.00632，其中变异度较大的是 *rps*3–*rpl*22、*trnL*–*UAC*–*rpl*32、*rpoB*–*trnC*–*GCA* 以及 *ycf*4，这些高变区可开发为分子标记用于评估飘带兜兰遗传多样性。(5) 系统发生分析结果表明，飘带兜兰 6 个个体叶绿体基因组序列聚在一起，与长瓣兜兰互为姐妹群。综上所述，飘带兜兰叶绿体基因组的 SSRs、长序列重复、SNPs、InDels 以及核苷酸序列呈现了足够的种内多样性，可开发成分子标记用于该种的系统演化及保护生物学研究。

关键词：叶绿体全基因组，飘带兜兰，序列差异，多态性分子标记，插入缺失，微卫星

中图分类号：Q943　　　　**文献标识码**：A

Intraspecific genetic variation within chloroplast genome of a rare and endangered species *Paphiopedilum parishii* (Orchidaceae)

GAO Xinzhen[1,2], TANG Lu[1], WANG Yu[1,3], SHAO Shicheng[1], LUO Yan[1*]

(1. *Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences,* Mengla, Yunnan 666303, China;
2. *University of Chinese Academy of Sciences,* Beijing 100049, China; 3. *School of Landscape Architecture, Southwest Forestry University,* Kunming 650224, China)

Abstract: *Paphiopedilum parishii* is a rare and endangered species with few localities and fragmented populations found in China, Myanmar, Thailand and Lao. Environmental changes and over-harvesting have led to a decrease in its wild populations. It is important to protect endangered

基金项目：国家自然科学基金 (31870183，32270225) [Supported by National Natural Science Foundation of China (31870183, 32270225)]。

作者简介：高鑫祯 (1997-)，硕士，研究方向为保护生物学，(E-mail) gaoxinzhen@xtbg.ac.cn。

***通信作者**：罗艳，博士，研究员，研究方向为兰科植物多样性与保护，(E-mail) luoyan@xtbg.org.cn。

species' genetic diversity since it provides the species with the ability to adapt and survive. However, little is known about the genetic information of this species. This study aims to detect intraspecific variation and develop polymorphic genetic markers of *P. parishii*. The complete chloroplast genome of four individuals of wild *P. parishii* was obtained by sequencing, assembling and annotating, then compared with the existing genomic data of two individuals available from GenBank to detect the intraspecific variation. Further, simple sequence repeats (SSRs), single nucleotide polymorphisms (SNPs), and insertions and deletions (InDels) were identified. The results were as follows: (1) The four new sequencing chloroplast genomes were quadripartite, with a length between 154 403 bp and 154 809 bp, with 129 genes (78 protein coding genes, 39 tRNAs, 8rRNAs and 4 pseudogenes). (2) As a result of comparison of six individuals, 103−107 SSR loci were identified in the chloroplast genome of six individuals of *P. parishii*, and 21SSRs were polymorphic. And 60 long repeats were found, including 17−21 forward repeats, 18−29 reverse repeats, 9−16 palindromic repeats, and 4−9 complement repeats. (3) In addition, a total of 10 SNPs and 60 InDels were uncovered across the plastome. Three of the non-synonymous mutations caused amino acid changes in functional domains. 19 InDels might be selected for possible chloroplast DNA markers to determine intraspecific variation. (4) The value of nucleotide diversity (P_i) was calculated ranging from 0−0.006 32 suggesting sequences with low variation. Hyper-polymorphic regions, e.g. intergenic spacers *rps*3−*rpl*22*, trnL−UAC−rpl*32*, rpoB−trnC−GCA* and *ycf*4 gene were identified as potential barcoding regions. (5) The phylogenetic analyses based on the complete chloroplast genome supported three lineages in *Paphiopedilum*, and 6 individuals of *P. parishii* form a monophyletic group. SSRs, long repeats, InDels, SNPs and nucleotide sequences showed sufficient intraspecific genetic variation in *P. parishii.* The molecular markers developed here will contribute to further evolutionary studies and conservation of *P. parishii*.

Key words: Chloroplast genome, *Paphiopedilum parishii*, sequence divergence, polymorphic DNA markers, InDels, SSRs

遗传多样性是生物多样性的基础，代表着物种适应环境和生存发展的能力，保存珍稀濒危物种的遗传多样性是物种保护的重要目标 (Guerrant & Pavlik, 1998; 黄宏文, 2018)。叶绿体是植物细胞内进行光合固碳和胁迫应答的重要半自主遗传细胞器，植物叶绿体基因组为环状 DNA 分子，单亲遗传，基因组大小为 107~218 kb，基因结构和组成稳定 (张韵洁和李德铢, 2011; Ivanova et al., 2017)。叶绿体基因组通常为四分体结构，包括两个反向重复序列 (inverted repeat, IR)、一个大单拷贝区 (large single copy region, LSC) 和一个小单拷贝区 (small single copy region, SSC) (Wolfe et al., 1987; 田欣和李德铢, 2002; Jansen et al., 2005; 张韵洁和李德铢, 2011; He et al., 2019)。叶绿体基因组编码 110~130 个基因，主要包括光合作用相关和叶绿体基因表达相关的基因等 (张韵洁和李德铢, 2011)。叶绿体基因组虽然较小，但包含有大量遗传信息，在核苷酸序列水平和结构重排上呈现出进化与保守性，编码区和非编码区有不同进化速率，在分子水平上有明显差异，可用于区分种间关系、评估种间变异性等，被广泛用于群体遗传多样性、系统演化等研究领域 (Wolfe et al., 1987; 张韵洁和李德铢, 2011; He et al., 2019; 黎若竹等, 2022)。近年来，利用种内不同个体叶绿体基因组的种内变异，为物种进化研究、种质资源利用以及濒危植物保护提供了基础遗传信息（Ishizuka et al., 2017; Muraguri et al., 2020; Zhang et al., 2020b）。

兜兰属 (*Paphiopedilum*) 为兰科杓兰亚科 (Orchidaceae, subfamily Cypripedioideae)植物，有 80 余种，主要分布于亚洲至太平洋岛屿的热带亚热带地区的石灰岩山地 (Cribb, 1998;

刘仲健等, 2009)。兜兰的唇瓣因形似拖鞋，被称为“拖鞋兰”，具有极高的观赏价值 (杨颖婕等, 2021)。近年来，由于兜兰属植物的自然生境局限于生态环境脆弱的喀斯特地区，人类活动导致的生态退化和生境破坏严重，加之过度采挖，使得兜兰属植物野生种群数量急剧减少，部分兜兰属植物面临野外灭绝的风险 (罗毅波等, 2003; 杨颖婕等, 2021)。目前，所有野生兜兰种类均被列入《濒危野生动植物种国际贸易公约》 (Convention on International Trade in Endangered Species of Wild Fauna and Flora, CITES) 禁止贸易，同时我国分布的野生兜兰属植物也全部列入《国家重点保护野生植物名录》(2021 年 9 月颁布实施)。我国野生兜兰属物种的濒危机制与保护工作急需开展。

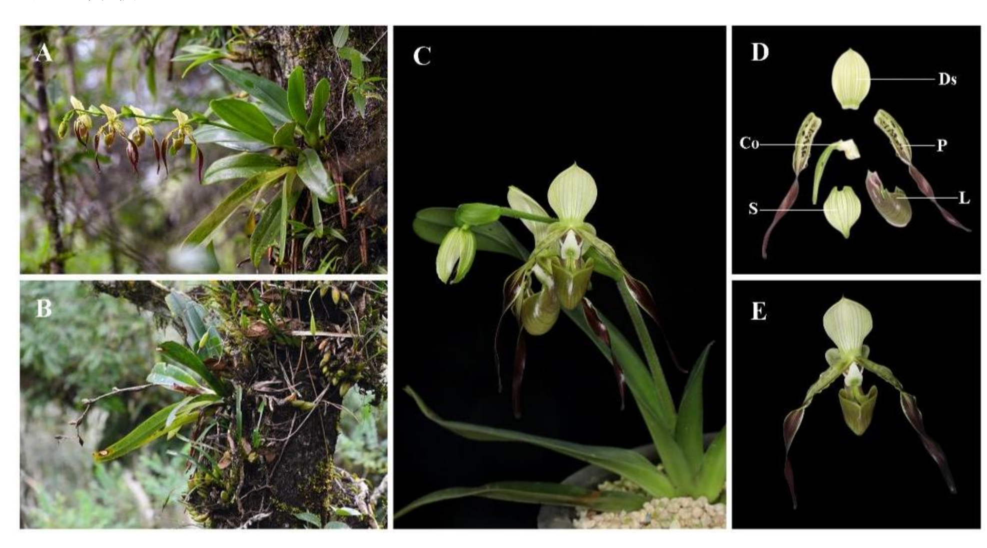
飘带兜兰 (*Paphiopedilum parishii*)为附生兰科植物，生于海拔 1 000~1 100 m 的树干或石头上，为我国一级重点保护植物 (图 1)。飘带兜兰被世界自然保护联盟 (International Union for Conservation of Nature, IUCN) 评估为濒危等级 (Endangered, ER)，仅在我国云南南部、泰国、缅甸和老挝分布有少数的野生种群 (Rankou & Averyanov, 2015)。飘带兜兰具有极高的观赏价值，在各植物园中均有引种，但野生种群已经极为罕见。Chen 等 (2012) 在云南南部进行了飘带兜兰繁殖生物学的观察，发现该种具有独特的自动授粉机制。基于叶绿体全基因组的发育基因组学的研究，飘带兜兰属于兜兰亚属 (subgenus *Paphiopedilum*) (Guo et al., 2021)。但目前有关飘带兜兰的遗传信息仍不完善，尚缺乏评估遗传多样性及种内多样性的分子标记，不利于对其开展综合保护。

本研究选取飘带兜兰的 4 个野生个体，进行浅层基因组测序，组装获得其叶绿体全基因组，并与已公开的该种 2 个个体叶绿体基因组 (Guo et al., 2021; Kao et al., 2021) 进行基因组比较研究，分析飘带兜兰叶绿体基因组的种内变异，理解飘带兜兰的基因进化水平和遗传多样性，以期开发种内多态性的分子标记用于保护生物学研究，为开展飘带兜兰的濒危机制及物种保护研究提供遗传信息基础资料。

1 材料与方法

1.1 研究材料及 DNA 提取、测序

飘带兜兰 4 个个体分别采自云南省勐腊县 (*P. parishii*_1 和 *P. parishii*_2) 及澜沧县 (*P. parishii*_3 和 *P. parishii*_4) 两个野生居群 (图 1)，取新鲜叶片放入硅胶中干燥。采用植物基因组 DNA 提取试剂盒 (TIANGEN, 中国北京) 提取叶片总 DNA，使用 IlluminaTruSeq 文库建立试剂盒 (San Diego, CA, USA) 构建测序文库，利用 Illumina Hiseq 2500 测序平台 (上海派森诺生物科技有限公司) 进行浅层基因组测序。同时，在 GenBank 上下载已公开的另外 2 个飘带兜兰个体的叶绿体基因组序列 (*P. parishii*_5, MW528213 和 *P. parishii*_6, MN587822) 进行比较基因组分析。



A. 生境；**B**. 结果植株；**C**. 开花植株；**D**. 花部结构图；**Ds**. 中萼片； **P**. 花瓣；**S**. 合萼片；**Co**. 合蕊柱；**L**. 唇瓣；**E**. 花。

A. Habitat; **B**. Fruiting individual; **C**. Flowering individual; **D**. Dissected flower; **Ds**. Dorsal sepal; **P**. Petals; **S**. Synsepals; **Co**. Column; **L**. Lip; **E**. Flower.

图 1 飘带兜兰

Fig. 1 *Paphiopedilum parishii*

1.2 叶绿体基因组的组装与注释

采用 Fastp 软件对下机后的原始数据 (raw data) 进行过滤，获得高质量测序数据 (clean data) 后，用 GetOrganelle v1.6.3 程序将测序后得到的数据组装成叶绿体基因组 (Jin et al., 2018)。组装完成后的叶绿体基因组数据用 Geneious Prime v2021.2.2 中的 MAFFT 工具进行比对后，使用在线工具 CPGAVAS2 (http://47.90.241.85:16019/analyzer/annotate2) (Shi et al., 2019)，GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) (Tillich et al., 2017) 和 tRNA-scan-SE v2.0.3 (Lowe & Chan, 2016) 对叶绿体基因组的蛋白质编码区、核糖体 RNA (rRNA) 和转运 RNA (tRNA) 进行注释，参考已发表的兜兰属叶绿体基因组对注释结果进行校正。最后通过 OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw. html) 在线网站绘制飘带兜兰叶绿体基因组图谱 (Lohse et al., 2007)。使用 Geneious Prime 2021.2.2 软件分析 6 个飘带兜兰叶绿体基因组的基本结构，统计 LSC、SSC 和 IR 的长度、GC 含量、蛋白质编码基因、tRNA 基因、rRNA 基因数目。

1.3 叶绿体基因组重复序列与 SSR 分析

通过 MISA 在线网站 (https://webblast.ipk-gatersleben.de/misa) (Beier et al., 2017) 测算叶绿体基因组中简单重复序列 (simple sequence repeats, SSR) 的分布，将核苷酸最小重复数分别设置为 10 (单核苷酸重复 mono-nucleotide)、5 (二核苷酸重复 di-nucleotide)、4 (三核苷酸重复 tri-nucleotide)、3 (四核苷酸重复 tetra-nucleotide)、3 (五核苷酸重复 penta-nucleotide)、3 (六核苷酸重复 hexa-nucleotide)。

通过 REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) (Kurtz et al., 2001) 在线网站测算叶绿体基因组中其他重复序列的分布，参数设置为最小重复序列 30 bp，汉明距离 (Hamming Distance) 设置为 3 (Liang et al., 2019)。

1.4 种间叶绿体基因组的比较分析

通过 mVISTA (https://genome.lbl.gov/vista) 在线网站 (Frazer et al., 2004)，以 *P.parishii*_1 的叶绿体基因组作为参考，将其余 5 条飘带兜兰的叶绿体基因组的编码区、基因间隔区、内含子区进行比对分析，评估各基因组序列之间的相似性和差异。使用 R v4.0.2 软件运行 IRscope (Amiryousefi et al., 2018) 脚本，对 6 个个体的叶绿体基因组序列的相邻边界区域进行绘制，比较分析边界区域的扩张和收缩情况。使用 DnaSP v5.10.0.1 软件，统计飘带兜兰不同个体间的序列的单核苷酸变异 (SNPs) 和插入缺失 (InDels) 情况，计算各序列间核苷酸变异值 (P_i)。

1.5 系统发育分析

从 GenBank 上下载 15 种兜兰属叶绿体全基因组序列，与本研究新测序拼装的 4 个个体以及已发表的 2 条飘带兜兰叶绿体基因组序列，以 3 种杓兰属植物为外类群来构建系统发育树 (物种的基因登录号见图 5)。使用 PhyloSuite v1.2.2 软件，利用 ModelFinder 计算最佳碱基替换模型 (Zhang et al., 2020a)，使用叶绿体基因组全序列，经 MAFFT 进行多序列比对后，采用 IQ-TREE 构建最大似然法 (maximum likelihood, ML) 系统发育树 (Trifinopoulos et al., 2016)，系统发育树分支的支持率基于 bootstrap 自展法检测，bootstrap 值设置为 5 000，重复次数设置为 1 000 次。

2 结果与分析

2.1 飘带兜兰叶绿体基因组基本特征

对原始数据进行过滤后，飘带兜兰 4 个样品分别获得 2.4~3.1 G 的高质量 clean data。组装和注释的叶绿体全基因组序列已登录至 GenBank (登录号分别为 OP 604356~OP 604359)。本研究将获得的飘带兜兰 4 个个体和已公布的 2 个飘带兜兰叶绿体基因组一起比对分析，6 条序列的叶绿体基因组均为环状四分体结构，包括 LSC 区，SSC 区和两个 IR 区 (图 2)。叶绿体基因组序列全长为 154 403~154 809 bp，其中 LSC 的长度为 86 581~86 983 bp，SSC 的长度为 2 436~2 446 bp，IR 的长度为 32 690~32 693 bp。总 GC 含量为 35.9%，LSC 的 GC 含量为 33.4%，SSC 的 GC 含量为 29.1%~29.2%，IR 的 GC 含量为 39.5% (表 1)。

飘带兜兰 6 个个体叶绿体基因组的基因总数均为 129 个，包括 78 个蛋白质编码基因、39 个 tRNA 基因、8 个 rRNA 基因，以及 4 个假基因 (*ndhJ*、*ndhD*、2 个 *ycf*15) (附录 1)。22 个基因在 IR 区重复，具有 2 个拷贝，包括 9 个蛋白质编码基因 (*psaC*、*ndhB*、*rps*7、*rps*15、*rps*19、*rpl*2、*rpl*23、*ycf*1、*ycf*2)、9 个 tRNA 基因 (*trnA-UGC*、*trnH-GUG*、*trnI-CAU*、*trnI-GAU*、*trnL-CAA*、*trnL-UAG*、*trnN-GUU*、*trnR-ACG*、*trnV-GAC*) 和 4 个 rRNA 基因 (*rrn*16、*rrn*23、*rrn*4.5、*rrn*5)。18 个基因有内含子，其中 15 个基因有 1 个内含子，包括 9 个蛋白编码基因 (*petB*、*petD*、*atpF*、*ndhB*、*rpoC*1、*rps*16、*rpl*2、*rpl*16、*accD*) 和 6 个 tRNA 基因 (*trnA-UGC*、*trnG-UCC*、*trnI-GAU*、*trnK-UUU*、*trnL-UAA*、*trnV-UAC*)；有 2 个基因含有 2 个内含子 (*clpP*、*ycf*3)；*rps*12 为反剪接基因 (附录 1)。

表 1 飘带兜兰 6 个个体叶绿体基因组基本特征

Table 1 The basic characteristics of the chloroplast genomes of six *Paphiopedilum parishii* individuals

基因组特征 Genome characteristics	*P. parishii*_1 (OP604356)	*P. parishii*_2 (OP604357)	*P. parishii*_3 (OP604358)	*P. parishii*_4 (OP604359)	*P. parishii*_5 (MW528213)	*P. parishii*_6 (MN587822)
序列全长 Genome size (bp)	154 809	154 660	154 809	154 403	154 689	154 692
LSC 区长度 LSC length (bp)	86 983	86 834	86 983	86 581	86 863	86 866
SSC 区长度 SSC length (bp)	2 446	2 446	2 446	2 436	2 446	2 446
IR 区长度 IR length (bp)	32 690	32 690	32 690	32 693	32 690	32 690
GC 含量 GC content (%)	35.9	35.9	35.9	35.9	35.9	35.9
LSC 区 GC 含量 GC content of LSC (%)	33.4	33.4	33.4	33.4	33.4	33.4
SSC 区 GC 含量 GC content of SSC (%)	29.1	29.2	29.1	29.2	29.2	29.2
IR 区 GC 含量 GC content of IR (%)	39.5	39.5	39.5	39.5	39.5	39.5
CDS 区 GC 含量 GC content of CDS (%)	37.5	37.5	37.5	37.5	37.5	37.5
tRNA 区 GC 含量 GC content of tRNA (%)	53.2	53.2	53.2	53.2	53.2	53.2
rRNA 区 GC 含量 GC content of rRNA (%)	55.2	55.2	55.2	55.2	55.2	55.2
基因总数 Total genes	129	129	129	129	129	129
假基因数 Pseudogene	4	4	4	4	4	4
编码蛋白基因数 No. of CDS	78	78	78	78	78	78
tRNA 基因数 No. of tRNA genes	39	39	39	39	39	39
rRNA 基因数 No. of rRNA genes	8	8	8	8	8	8

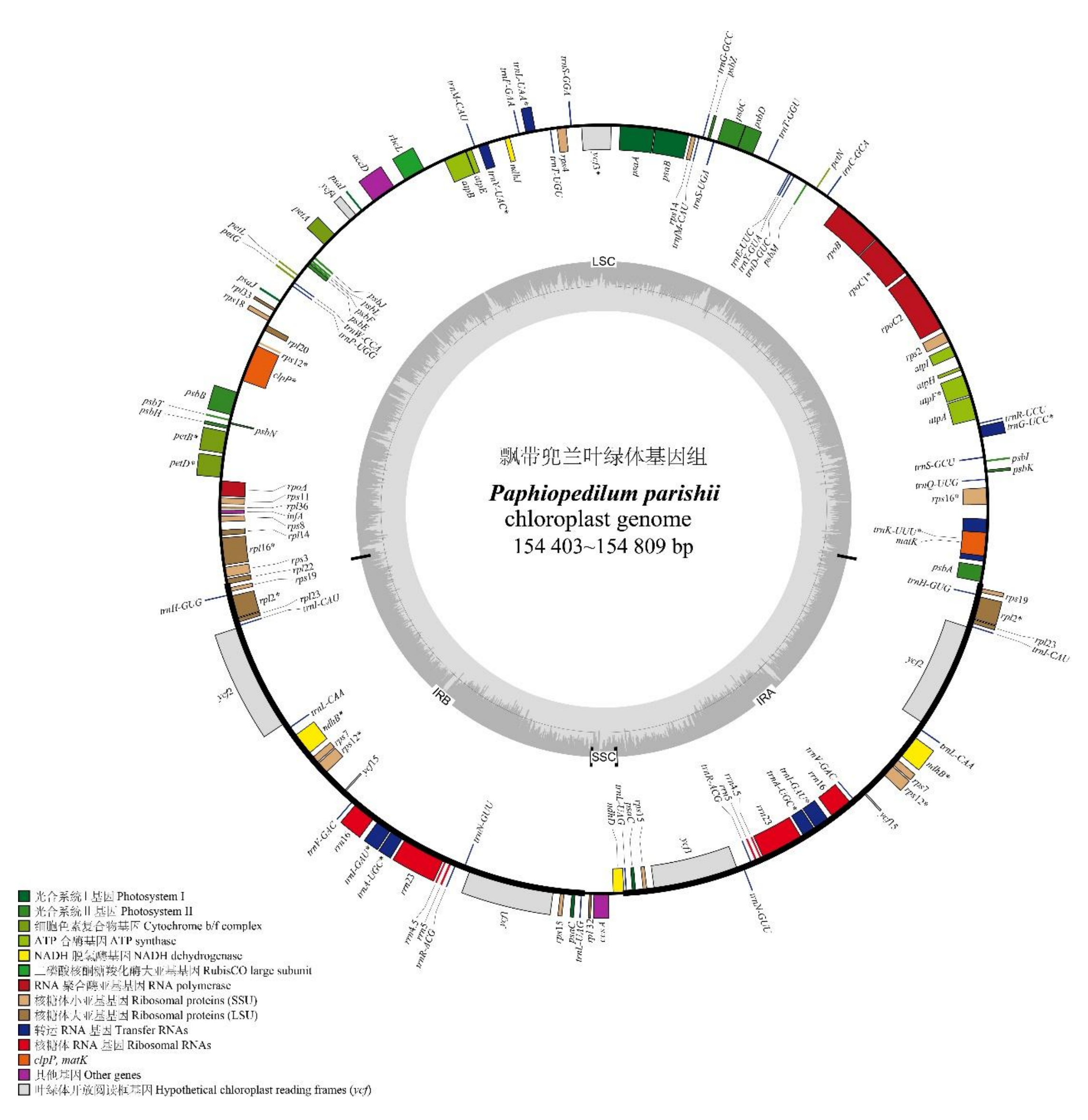


图 2 飘带兜兰叶绿体基因组图谱

Fig.2 Chloroplast genome map of *Paphiopedilum parishii*

2.2 SSR 位点分析

利用 MISA 软件在飘带兜兰 6 个体的叶绿体基因组中分别检测到 103~107 个 SSR 位点，包括 47~51 个单核苷酸重复（主要以 A, T 碱基重复为主）、20 个二核苷酸重复、15 个三核苷酸重复、14 个四核苷酸重复、2 个五核苷酸重复以及 5 个六核苷酸重复类型（图 3: A，附录 2）。SSR 位点大部分位于 LSC 区（84~88 个），其次 IR 区分布有 18 ~19 个，SSC 区无 SSR 位点（图 3: B）。另外，SSR 位点主要分布于叶绿体基因组非编码区的基因间隔区（Intergenic Spacer, IGS），为 72 ~ 76 个；位于基因内含子区（Intron）的有 17 个；位于编码区的有 14 个（图 3: C, 3: D）。

将飘带兜兰 6 个个体同源区域的 SSR 位点进行详细的比较，筛选出具有相同类型但重复数目不同的 SSR 位点，可作为种内多态性位点，共发现 21 个 SSR 位点在 6 个个体中有差异（表 2）。

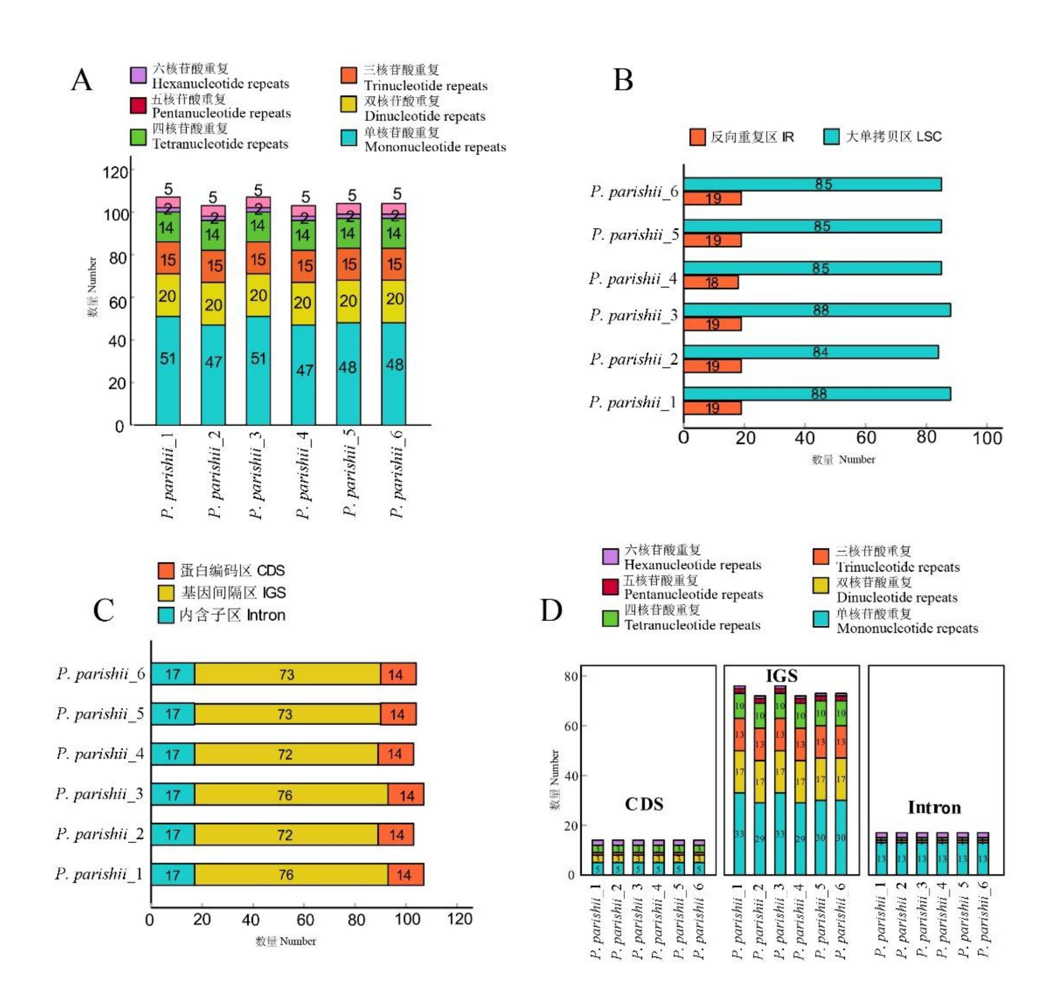
表 2 飘带兜兰 6 个个体中存在的多态性 SSR 位点

Table 2 The polymorphic simple sequence repeats (SSR) in the chloroplast genomes of six *Paphiopedilum parishii* individuals

SSR 类型 SSR type	*P. parishii*_1/*P. parishii*_2/*P. parishii*_3/ *P. parishii*_4/*P. parishii*_5/*P. parishii*_6	位置 Location	区域 Region
A	0/0/0/10/0/0	*trnK-UUU*	Intron
A	11/0/11/0/0/0	*trnK-UUU-rps*16	IGS
T	17/15/17/20/15/15	*atpF*	Intron
T	10/0/10/10/0/0	*atpH-atpI*	IGS
A	10/11/10/0/11/11	*atpI-rps*2	IGS
A	0/0/0/10/0/0/0	*petN-psbM*	IGS
T	10/0/10/0/0/0	*trnL-UAA-trnF-GAA*	IGS
T	10/0/10/0/0/0	*trnF-GAA-ndhJ*	IGS
T	12/10/12/14/10/10	*atpB-rbcL*	IGS
A	0/10/0/10/10/10	*rbcL-accD*	IGS
T	0/10/0/10/10/10	*petA-psbJ*	IGS
T	15/14/15/10/19/20	*trnP-UGG-psaJ*	IGS
T	23/0/23/24/24/24	*trnP-UGG-psaJ*	IGS
T	10/10/10/0/10/10	*clpP*	Intron
T	12/12/12/0/12/12	*clpP*	Intron
T	10/0/10/0/0/0	*rpl*36-*infA*	IGS
T	15/12/15/10/12/12	*rpl*14-*rpl*16	IGS
G	10/10/10/0/10/10	*trnI-GAU*	Intron
TA	8/8/8/6/8/8	*trnE-UUC-trnT-GGU*	IGS
TA	9/9/9/11/9/9	*psbB-psbT*	IGS
GGAAGA	6/6/6/5/6/6	*ycf*1	CDS

注：数字代表重复个数。

Note: Number indicated SSR repeats.



A. 不同类型重复单元的数量；B. SSR位点在叶绿体基因组大单拷贝区（LSC），小单拷贝区（SSC）和反向重复区（IR）的数量；C. SSR位点在CDS，IGS，Intron区域的数量；D. SSR位点分别在叶绿体基因组CDS，IGS，Intron区域不同重复类型的数量分布。

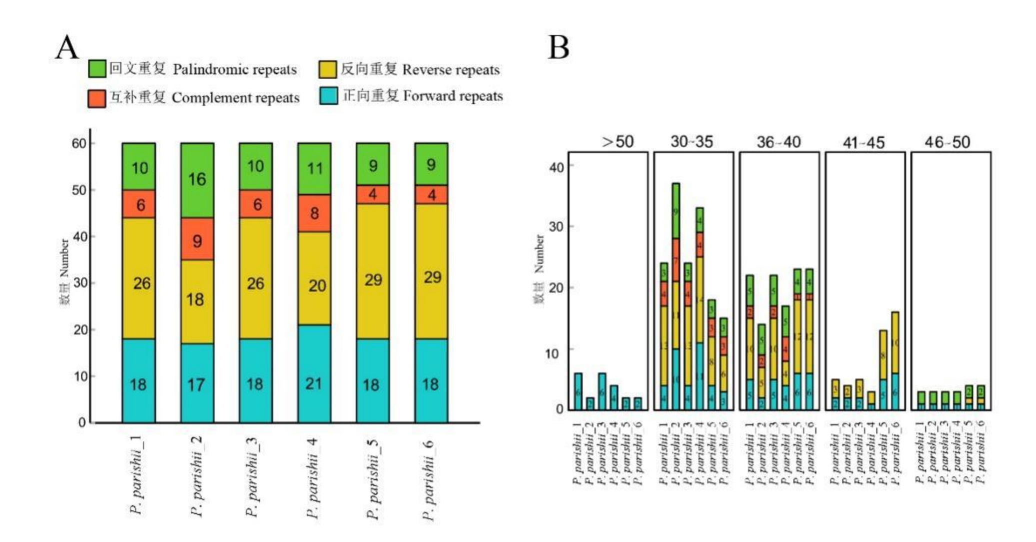
A. Numbers of SSRs in different types; B. Numbers of SSRs in the LSC (large single copy), SSC (small single copy), and IR (invert repeat) regions; C. Numbers of SSRs identified in CDS (coding regions), IGS (intergenic spacers), and introns of the cp genome; D. Frequency distribution of different types of SSRs identified in CDS, IGS, and introns of the cp genome.

图 3 飘带兜兰叶绿体基因组 SSR 分析

Fig.3 SSR analysis of the chloroplast genomes of *Paphiopedilum parishii*

2.3 重复序列分析

利用 REPuter 对 6 条飘带兜兰叶绿体基因组序列进行重复序列分析。在 6 个叶绿体基因组中均鉴定到 60 个重复序列，包括 17~21 个正向重复（forward repeats)；18~29 个反向重复(reverse repeats)；4~9 个互补重复（complement repeats）以及 9~16 个回文重复（palindromic repeats) (图 4: A)。6 条叶绿体基因组重复序列的长度范围跨度较大，但都集中在 30~40 bp 之间（图 4: B)。



A. 不同类型重复序列的数量；**B.** 重复序列的类型在不同重复长度中的数量分布。

A. Numbers of repeat sequences in different types; **B.** Frequency distribution of different types of repeat sequence identified in different sizes of the chloroplast.

图4 飘带兜兰叶绿体基因组重复序列分析

Fig.4 Repeat sequences analysis of the chloroplast of *Paphiopedilum parishii*

2.4 飘带兜兰种间叶绿体基因组序列的多样性

以 *P. parishii*_1 为参考序列，利用 mVISTA 软件对 6 条飘带兜兰序列进行序列多样性比较（附录 3)。结果表明，6 条飘带兜兰序列相似性较高，LSC 区的多样性相对较高，略高于 IR 区和 SSC 区。多样性水平较高的片段主要在非编码区，例如 *atpH* - *atpI*、*psaA* - *ycf*3、*atpB* - *rbcL*、*accD* - *psaI*、*trnP-UGG* - *psaJ*。而编码区变异度较低，*ycf*1 基因的变异度稍高，tRNA 和 rRNA 基因在飘带兜兰中极为保守。

6 条序列边界比较图显示，*rpl*22、*trnL-UAG*、*rps*19 以及 *psbA* 分别位于 6 条序列的 LSC/IRb、IR/SSC、IRa/LSC 边界或边界附近。*rpl*22 位于 LSC/IRb 边界上，有 58 bp 的片段位于 IRb 区域；重复基因 *trnL-UAG* 位于 IR 区域内，离 IR/SSC 边界相距 176~186 bp；重复基因 *rps*19 位于 IR 区，离 IR/LSC 边界 282 bp；*psbA* 位于 LSC 区，离 IRa/LSC 边界 96 bp （附录 4)，6 条序列边界基因较为保守，扩张收缩现象不显著，只有 *P. parishii_4* 的 IR 区有微小扩张。

2.5 飘带兜兰种间叶绿体基因组核苷酸多样性

通过对 6 条飘带兜兰叶绿体基因组核苷酸多样性的分析，发现 6 条叶绿体基因组序列中共存在 10 个 SNPs 以及 60 个 InDels，其中基因间隔区有 4 个 SNPs，43 个 InDels，内含子区有 1 个 SNP 和 13 个 InDels，编码区有 5 个 SNPs 和 4 个 InDels (表 3)。在 10 个 SNPs 中有 3 个发生了非同义替换，导致密码子改变，分别是在 *rpoC*1 基因中碱基 A 替换为 C，使编码的氨基酸由异亮氨酸转变为丝氨酸；在 *rpoB* 基因中碱基 C 替换为碱基 T，使编码的甘氨酸转变为精氨酸；在 *ycf*4 基因中碱基 G 替换为 C，使编码的蛋氨酸转变为异亮氨酸。此外还有 2 个 SNPs 发生同义替换，密码子并未改变，其余 SNPs 均位于非编码区。我们还对大于 3 bp 的 InDels 进行了多态性的分析，发现共有 19 个位点呈现了种内的多态性 (表 4)，例如基因间隔区 *pasA* - *ycf*3 分别在 *P. parishii*_1、3、4 中发生了插入，而在 *P. parishii*_2、5、6 中缺失；而在 *accD* 基因的编码区及内含子分别存在的 2 个 InDels 在个体间也具有多态性。

利用 DnaSP 分别计算了飘带兜兰 6 个个体的叶绿体基因组基因和基因间隔区的核苷酸多样性值（P_i），编码区的 P_i 值为 0~0.000 61，仅有 5 个基因的核苷酸多样性值大于 0，其中 *ycf*4 基因的 P_i 值最高；基因间隔区的 P_i 值为 0~0.006 32，仅有 3 个基因间隔区的核苷酸多样性值大于 0，其中 *rps*3 - *rpl*22 的 P_i 值最高（附录 5），表明飘带兜兰种内表现了较低水平的核苷酸多样性，基因间隔区的核苷酸多样性相对较高。此外，分析发现变异程度较高的位点主要位于 LSC 区，而 IR 区和 SSC 区相对保守。

表 3 飘带兜兰叶绿体基因组插入缺失统计

Table 3 Summary of variants detected in *Paphiopedilum parishii* chloroplast genomes

	插入缺失 InDels			单核苷酸变异 SNPs		
总数 Total	蛋白编码区 CDS	基因间隔区 IGS	内含子区 Intron	蛋白编码区 CDS	基因间隔区 IGS	内含子区 Intron
70	4	43	13	5	4	1

注：CDS．编码序列；IGS．基因间隔区；InDels．插入缺失；SNP．单核苷酸多态性。

Note: CDS. Coding sequence; IGS. Intergenic spacer; InDels. Insertion or deletion; SNP. Single nucleotide polymorphism.

表 4 飘带兜兰叶绿体基因组中的插入缺失（>3 bp）

Table 4 Characters of InDels (>3 bp) in chloroplast genomes of six *Paphiopedilum parishii* individuals

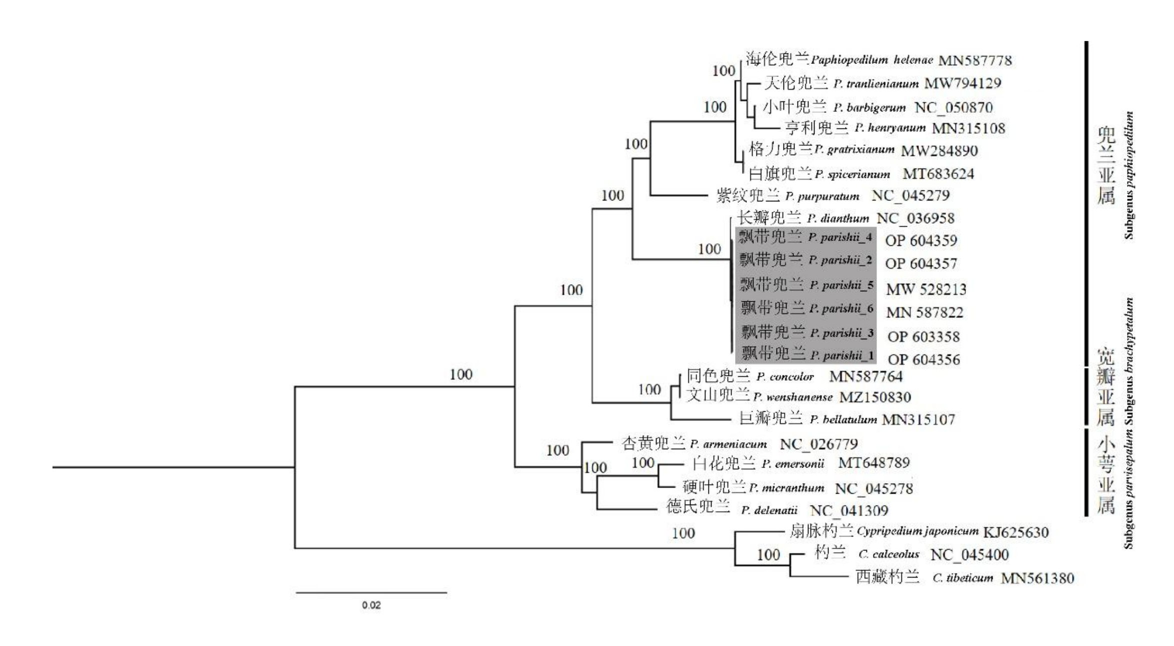
插入缺失大小 InDels size (bp)	*P. parishii*_1/*P. parishii*_2/*P. parishii*_3/ *P. parishii*_4/*P. parishii*_5/*P. parishii*_6	位置 Location	区域 Region
18	0/1/0/1/1/1	*trnS-GCU* - *trnG-UCC*	IGS
5	0/0/0/1/0/0	*atpF*	Intron
273	1/1/1/0/1/1	*atpH* - *atpI*	IGS
4	1/1/1/0/1/1	*trnE-UUC* - *trnT-GGU*	IGS
19	1/0/1/1/0/0	*psaA* - *ycf*3	IGS
7	1/0/1/0/0/0	*trnT-UGU* - *trnL-UAA*	IGS
4	1/1/1/0/1/1	*ndhJ*	CDS*
4	0/0/0/1/0/0	*atpB* - *rbcL*	IGS
7	0/1/0/0/1/1	*atpB* - *rbcL*	IGS
12	1/1/1/0/1/1	*accD*	CDS
54	1/0/1/0/0/0	*accD*	Intron
89	1/1/1/0/1/1	*accD* - *psaI*	IGS
31	1/0/1/0/1/1	*trnP-UGG* - *psaJ*	IGS
60	1/0/1/1/0/0	*trnP-UGG* - *psaJ*	IGS
4	0/0/0/1/0/0	*psbB* - *psbT*	IGS
5	1/0/1/0/0/0	*rpl*14 - *rpl*16	IGS
6	1/1/1/0/1/1	*ycf*1	CDS
8	0/0/0/1/0/0	*ccsA* - *trnL-UAG*	IGS
6	1/1/1/0/1/1	*ycf*1	CDS

注：1 表示插入；0 表示缺失。

Note：1 indicates insersion; 0 indicates deletion.

2.6 飘带兜兰的系统发育关系

通过 PhyloSuite 计算出最佳碱基替换模型为 TVM+F+R2，采用此模型构建系统发育树。16 种兜兰属植物构成了支持率很高（支持率均为 100%）的 3 个分支，其中德氏兜兰（*P. delenatii*）、杏黄兜兰（*P. armeniacum*）、白花兜兰（*P. emersonii*）、和硬叶兜兰（*P. micranthum*）构成了基部分支，归属于小萼亚属（Subgenus *Parvisepalum*）；文山兜兰（*P. wenshanense*）、同色兜兰（*P. concolor*）和巨瓣兜兰（*P. bellatulum*）形成一个单系分支，归属于宽瓣亚属（Subgenus *Brachypetalum*），本研究中的飘带兜兰 6 个个体并未分开，形成 1 个分支，与长瓣兜兰形成姐妹群（图 5）。



分支上的数值代表自展值。

The numbers above branches represent bootstrap percentage (BP) of ML.

图 5 基于叶绿体全基因组序列最大似然法构建的系统发育树

Fig. 5 Phylogenetic tree conducted using Maximum Likelihood (ML) methods based on whole chloroplast genomes from different species

3 讨论与结论

本研究利用二代高通量测序技术对飘带兜兰野生居群不同个体进行浅层基因组测序，通过组装和注释获得到了飘带兜兰叶绿体全基因组。6 个个体的叶绿体基因组的基因数量和顺序一致，没有发生基因组的重排或倒位事件。6 个叶绿体基因组均为叶绿体基因组的典型四分体式结构，由 IR 区、LSC 区、SSC 区组成（Jansen et al., 2005）。飘带兜兰的叶绿体基因组长度为 154 403~154 809 bp，符合此前报道的兰科及兜兰属植物的叶绿体基因组大小（Kim et al., 2014; Guo et al., 2021）。飘带兜兰不同个体间的叶绿体基因组长度变化主要在 LSC 区，为 86 581~86 983 bp，而 SSC 区和 IR 区基本保持一致，分别为 2 436~2 446 bp 和 32 690~32 693 bp。IR 区较其他被子植物有显著的扩张，SSC 区则有较大程度的收缩，仅包含 *rpl*32 和 *ccsA* 2 个基因，这些特点符合兜兰属叶绿体基因组特征（Guo et al., 2021）。与其他已报道的兰科属种的叶绿体全基因组的 SSC 区长度相比，例如凌氏石豆兰（*Bulbophyllum lingii*, 18 244 bp）、王氏石斛（*Dendrobium wangliangii*, 18 373 bp）、扇脉杓兰（*Cypripedium japonicum*, 21 911 bp) (Kim et al., 2020; Shao et al., 2020; Tang et al., 2021)，兜兰属的 SSC 区收缩严重，导致整个基因组的长度远低于兰科植物其他物种。由于大量 SSC 区的基因转移至 IR 区，使

得 SSC 区的收缩在兜兰属极为常见，杏黄兜兰中甚至一些典型的 SSC 区基因，如 *ycf*1、*psaC*、*ndhD* 转移至 IR 区 (Kim et al., 2015; Lin et al., 2015; Niu et al., 2017)。有研究认为由于 IR 区的基因可以利用同源重组机制修复 DNA 损伤，较大的 IR 区域更有利于质体基因组的稳定性，因而基因的转移可能更利于该基因的表达 (Palmer & Thompson, 1982; Wicke et al., 2011)。这种基因转移使得兜兰属 SSC 区具有极高多样性，序列长度差异大，长度为 524~5 913 bp，基因数目也显著不同，其种间多样性和种内的稳定性使其具有开发成用于物种鉴定的分子标记的潜力 (Guo et al., 2021)。

飘带兜兰的叶绿体基因组共编码 129 个基因，包括 78 个蛋白质编码基因、39 个 tRNA 基因、8 个 rRNA 基因，以及 4 个假基因，与此前报道的相同 (Guo et al., 2021)。4 个假基因都是由于大量基因片段缺失造成功能丧失，其中 *ndhJ* 由于片段缺失导致基因缺少起始密码子，而 *ndhD* 和 *ycf*15 的碱基缺失 50%以上。飘带兜兰叶绿体基因组中存在大量的 *ndh* 基因丢失，只保留了 *ndhB*、*ndhD* 和 *ndhJ*，其中 *ndhD* 和 *ndhJ* 为假基因，这也是 SSC 区收缩的原因之一。兰科植物普遍存在 *ndh* 基因的丢失现象 (Yang et al., 2013; Feng et al., 2016; Niu et al., 2017; Zavala-Páez et al., 2020)，但 *ndh* 基因的丢失程度不同，比如石豆兰属叶绿体基因组中仍存在 12 个 *ndh* 基因 (Tang et al., 2021)，兰属植物叶绿体基因组中存在 10 个 *ndh* 基因 (胡国家, 2020)，石斛属的 *ndh* 基因丢失现象极为严重 (牛志韬, 2017)。叶绿体 *ndh* 基因与核基因组部分基因共同编码 NAD (P) H 脱氢酶复合体，参与环式电子传递链 (CET) 途径，在植物抗逆胁迫中起重要作用。除了兰科植物外，在裸子植物松杉类和麻黄类中也发现 *ndh* 基因的丢失 (Braukmann et al., 2009; Wu et al., 2009)。有研究认为 *ndh* 基因的缺失可能是植物由自养转为异养的进化过程 (Lin et al., 2017)。

通过比较基因组学分析，本研究在飘带兜兰 6 个个体的基因组中共鉴定到 103~107 个 SSR 位点，其中数量最多的是单核苷酸重复序列，其次是二核苷酸重复序列，多以 A、T 碱基为基本重复单元，显示了高度的 A/T 偏好，与当前在被子植物叶绿体基因组中观察到的情况一致 (Qin et al., 2015; 陈模舜和杨仲毅, 2022)。重复序列在植物基因组中扮演重要角色，由于多态性较高，在群体遗传和进化研究中经常被用来做分子标记 (Muraguri et al., 2020)。本研究共筛选到 21 个 SSR 位点具有多态性，这些位点可开发为分子标记用于评估飘带兜兰种群的遗传多样性。此外，本研究还发现非编码区的 SSR 数量远多于编码区，这说明非编码区比编码区具有更高遗传多样性，这可能由于非编码区面临着更大的选择压力 (Shaw et al., 2007)。正向重复、反向重复、互补重复和回文重复在 6 个个体的叶绿体基因中也表现了多样性。同种植物长序列重复的不同可能由于序列的插入缺失导致重复序列的类型发生改变，可能与基因重组有关 (Somaratne et al., 2019)。

插入缺失和突变引起基因结构的差异是遗传变异的重要来源，代表生物个体适应环境变化的能力 (Han & Xue, 2003)。种间或种内的叶绿体基因结构差异可用于近缘种间和种内的系统发育关系及群体遗传学研究 (McCauley, 1995; Kersten et al., 2016)。本研究在飘带兜兰的 6 个个体的叶绿体基因组中鉴定到 60 个 InDels 和 10 个 SNPs，主要分布在基因间隔区。与其他种内叶绿体基因组比较分析相比，飘带兜兰的 InDels 和 SNPs 数量并不多。Muraguri 等 (2020) 在蓖麻 (*Ricinus communis*) 12 个个体的叶绿体基因组中共鉴定到 162 个 SNPs 和 92 个 InDels；Zhang 等 (2020b) 在麻栎 (*Quercus acutissima*) 3 个个体中共检测到 77 个 SNPs 和 255 个 InDels。Alexander 等 (2014) 在红槲栎 (*Q. rubra*) 4 个个体中鉴定到 8 个 SNPs 和 45 个 InDels。本研究共发现 19 个位点呈现种内多态性。通过 mVISTA 基因组比对和核苷酸多样性 (P_i) 值的计算，发现非编码区的核苷酸多样性远高于编码区，其中 *rps*3-*rpl*22 的 P_i 值最高 (0.006 32)。飘带兜兰叶绿体基因组中存在的这些基因组的结构变异可广泛用于种间和种内关系以及遗传多样性研究。

遗传多样性是评估物种对栖息地环境适应能力以及抗逆能力的重要指标，研究遗传多样

性和遗传结构是制定物种保护措施的前提（张亚红等, 2019)。通过叶绿体全基因组的比较分析，本研究筛选出了一些多态性叶绿体基因组分子标记，为评估飘带兜兰的遗传多样性，探讨系统发生、进化以及保护遗传学研究提供遗传学基础。

致谢 中国科学院西双版纳热带植物园 Sven Landrein、杨国平协助野外调查，云南野兰堂生物有限公司张泽提供照片，谨致谢意。

参考文献：

ALEXANDER LW, WOESTE KE, 2014. Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management [J]. Tree Genet Genomes, 10: 803 - 812.

AMIRYOUSEFI A, HYVONEN J, POCZAI P, 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes [J]. Bioinformatics, 34(17): 3030 - 3031.

BEIER S, THIEL T, MUNCH T, et al., 2017. MISA-web: a web server for microsatellite prediction [J]. Bioinformatics, 33(16): 2583 - 2585.

BRAUKMANN TW, KUZMINA M, STEFANOVIC S, 2009. Loss of all plastid *ndh* genes in gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny [J]. Curr Genet, 55(3): 323 - 337.

CHEN LJ, LIU KW, XIAO XJ, et al., 2012. The anther steps onto the stigma for self-fertilization in a slipper Orchid [J]. PLOS ONE, 7(5): e37478.

CHEN MS, YANG ZY, 2022. Genealogical structure and differentiation analysis of *Carpinus tientaiensis* based on single nucleotide polymorphism of chloroplast genome [J]. Guihaia, 42(10): 1703 - 1716.[陈模舜，杨仲毅, 2022. 基于叶绿体基因组 SNP 的天台鹅耳枥谱系结构与分化分析 [J]. 广西植物, 42(10): 1703 - 1716.]

CRIBB P, 1998. The Genus *Paphiopedilum* [M]. Borneo, Malaysia: National History Publications.

FENG YL, WICKE S, LI JW, et al., 2016. Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species [J]. Genome Biol Evol, 8(7): 2164 - 2175.

FRAZER KA, PACHTER L, POLIAKOV A, et al., 2004. VISTA: computational tools for comparative genomics [J]. Nucl Acid Res, 32: W273 - W279.

GUERRANT EO, PAVLIK BM, 1998. Reintroduction of rare plants: Genetics, Demography, and the role of ex situ conservation methods[M]// FIEDLER PL, KAREIVA PM, Conservation Biology: For the Coming Decade. Boston, MA: Springer US: 80 - 108.

GUO YY, YANG JX, BAI MZ, et al., 2021. The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions [J]. BMC Plant Biol, 21(1): 248.

HAN B, XUE YB, 2003. Genome-wide intraspecific DNA-sequence variations in rice [J]. Curr Opin Plant Biol, 6(2):134 - 138.

HE J, YAO M, LYU RD, et al., 2019. Structural variation of the complete chloroplast genome and plastid phylogenomics of the genus *Asteropyrum* (Ranunculaceae) [J]. Sci Rep, 9(1):15285.

HU GJ, 2020. The complete chloroplast genomes of *Paphiopedilum* and *Cymbidium* (Orchidaceae) species: comparative genomic and phylogenetic analyses [D]. Xi’an: Northwest University: 28 -30. [胡国家, 2020. 兰科兜兰属与兰属植物比较叶绿体基因组学与系统发育研究 [D]. 西安：西北大学 28 - 30.]

HUANG HW, 2018. The principle and practice of ex situ plant conservation [M]. Beijing :

Science Press: 176 - 179.[黄宏文，2018. 植物迁地保育原理与实践 [M]. 北京: 科学出版社: 176 -179.]
ISHIZUKA W, TABATA A, ONO K, et al., 2017. Draft chloroplast genome of *Larix gmelinii* var. *japonica*: insight into intraspecific divergence [J]. J For Res, 22(6): 393 - 398.
IVANOVA Z, SABLOK G, DASKALOVA E, et al., 2017. Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response [J]. Front Plant Sci, 8: 204.
JANSEN RK, RAUBESON LA, BOORE JL, et al., 2005. Methods for obtaining and analyzing whole chloroplast genome sequences [M]//ZIMMER EA, ROALSON EH. Molecular Evolution: Producing the Biochemical Date: Part B: 348 - 384.
JIN J, YU WB, YANG JB, et al., 2018. GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data [J]. Genome Biol, 21(1): 241.
KAO H, ZHAO Y, YANG M, et al., 2021. The complete chloroplast genome sequences of an endangered orchid species *Paphiopedilum parishii* (Orchidaceae) [J]. Mitochondrial DNA B, 6(9): 2521 - 2522.
KERSTEN B, RAMPANT PF, MADER M, et al., 2016. Genome sequences of *Populus tremula* chloroplast and mitochondrion: Implications for holistic poplar breeding [J]. PLoS ONE, 11(1): e0147209.
KIM HT, KIM JS, MOORE M, et al., 2015. Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries [J]. PLoS ONE, 10(11): e0142215.
KIM JS, KIM HT, KIM JH, et al., 2014. The largest plastid genome of monocots: a novel genome type containing AT residue repeats in the slipper orchid *Cypripedium japonicum* [J]. Plant Mol Biol Rep, 33(5): 1 - 11.
KIM YK, JO SJ, SE-HWAN C, et al., 2020. Plastome evolution and phylogeny of Orchidaceae, with 24 new sequences [J]. Front Plant Sci, 11: 22.
KURTZ S, CHOUDHURI JV, OHLEBUSCH E, et al., 2001. REPuter: the manifold applications of repeat analysis on a genomic scale [J]. Nucl Acid Res, 29(22): 4633 - 4642.
LIANG CL, WANG L, LEI J, et al., 2019. A comparative analysis of the chloroplast genomes of four *Salvia* medicinal plants [J]. Engineering, 5(5): 907 - 915.
LIN CS, CHEN JJ, HUANG YT, et al., 2015. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family [J]. Sci Rep, 5(1): 9040.
LIN CS, CHEN JJW, CHIU CC, et al., 2017. Concomitant loss of *NDH* complex-related genes within chloroplast and nuclear genomes in some orchids [J]. Plant J, 90(5): 994 - 1006.
LIU ZJ, CHEN XQ, CHEN LJ, 2009. The genus *Paphiopedilum* in China [M]. Beijing : Science Press: 8 - 10. [刘仲健，陈心启，陈利君，2009. 中国兜兰属植物 [M]. 北京: 科学出版社: 8 - 10.]
LI RZ, CAI J, YANG JB, et al., 2022. Plastid phylogenomics resolving phylogenetic placement and genera phylogeny of Sterculioideae (Malvaceae s.l.) [J]. Guihaia, 2022,42(1): 25 - 38. [黎若竹，蔡杰，杨俊波，等，2022. 利用叶绿体基因组数据解析锦葵科梧桐亚科的系统位置和属间关系 [J]. 广西植物，42(1): 25 - 38.]
LOHSE M, DRECHSEL O, BOCK R, 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial

genomes [J]. Curr Genet, 52(5/6): 267 - 274.
LOWE TM, CHAN PP, 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes [J]. Nucl Acid Res, 44(1): W54 - W57.
LUO YB, JIA JS, WANG CL, 2003. A general review of the conservation status of Chinese orchids [J]. Biodivers Sci, 11(1): 70 - 77. [罗毅波，贾建生，王春玲，2003. 中国兰科植物保育的现状和展望 [J]. 生物多样性，11(1): 70 -77.]
MCCAULEY DE, 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants [J]. Trend Ecol Evol, 10(5):198 - 202.
MURAGURI S, XU W, CHAPMAN M, et al., 2020. Intraspecific variation within Castor bean (*Ricinus communis* L.) based on chloroplast genomes [J]. Ind Crops Products, 155: 112779.
NIU ZT, XUE QY, ZHU SY, et al., 2017. The complete plastome sequences of four orchid species: Insights into the evolution of the Orchidaceae and the utility of plastomic mutational hotspots [J]. Front Plant Sci, 8: 715.
NIU ZT, 2017. Comparative plastomic studies of *Dendrobium* species and comparison of the physiogical effects and transcriptome responses of *Dendrobium officinale* under different abiotic stresses [D]. Nanjing: Nanjing Normal University: 25 - 26. [牛志韬, 2017. 石斛属植物叶绿体基因组研究及其代表种逆境下光合作用途径探讨 [D]. 南京: 南京师范大学: 25 - 26.]
PALMER JD, THOMPSON WF, 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost [J]. Cell, 29(2): 537 - 550.
QIN Z, WANG YP, WANG QM, et al., 2015. Evolution analysis of simple sequence repeats in plant genome [J]. PLoS ONE, 10(12): e0144108.
RANKOU H, AVERYANOV L, 2015. *Paphiopedilum parishii*[J]. The IUCN red list of threatened species, 2015: e. T193512A2240580.
SHAO SC, TANG L, LUO Y, 2020. The complete chloroplast genome sequence of *Dendrobium wangliangii* (Orchidaceae) [J]. Mitochondrial DNA B, 5(3): 3513 - 3515.
SHAW J, LICKEY EB, SCHILLING EE, et al., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III [J]. Am J Bot, 94(3): 275 - 288.
SHI LC, CHEN HM, JIANG M, et al., 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer [J]. Nucl Acid Res, 47(1): W65 - W73.
SOMARATNE Y, GUAN DL, WANG WQ, et al., 2019. Complete chloroplast genome sequence of *Xanthium sibiricum* provides useful DNA barcodes for future species identification and phylogeny [J]. Plant Syst Evol, 305(10): 949 - 960.
TANG HQ, TANG L, SHAO SC, et al., 2021. Chloroplast genomic diversity in *Bulbophyllum* section *Macrocaulia* (Orchidaceae, Epidendroideae, Malaxideae): insights into species divergence and adaptive evolution [J]. Plant Diverers, 43(5): 350 - 361.
TIAN X, LI DZ, 2002. Application of DNA sequences in plant phylogenetic study [J]. Acta Botanica Yunnan, 24(2):170 - 184. [田欣，李德铢，2002. DNA 序列在植物系统学研究中的应用 [J]. 云南植物研究: 24(2):170 - 184.]
TILLICH M, LEHWARK P, PELLIZZER T, et al., 2017. GeSeq-versatile and accurate annotation of organelle genomes [J]. Nucl Acid Res, 45(1): W6 - W11.
TRIFINOPOULOS J, LAMTUNG N, VON HAESELER A, et al., 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis [J]. Nucl Acid Res, 44(1): W232 -

W235.
WICKE S, SCHNEEWEISS G, DEPAMPHILIS C, et al., 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function [J]. Plant Mol Biol, 76(3 - 5): 273 - 297.
WOLFE KH, LI WH, SHARP PM, 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs [J]. Proc Natl Acad Sci USA, 84(24): 9054 - 9058.
WU CS, LAI YT, LIN CP, et al., 2009. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: Selection toward a lower-cost strategy [J]. Mol Phylogenet Evol, 52(1): 115 - 124.
YANG JB, TANG M, LI HT, et al., 2013. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses [J]. BMC Evol Biol, 13(1): 84.
YANG YJ, HUANG JL, HU H, et al., 2021. Progress on conservation and utilization of *Paphiopedilum* species in China [J]. J W Chin For Sci, 50(5): 108 - 112. [杨颖婕，黄家林，胡虹，等，2021. 中国兜兰属植物种质资源保护和利用研究进展 [J]. 西部林业科学，50(5): 108 - 112 .]
ZAVALAPAEZ M, VIEIRA LDN, DEBAURA VA, et al., 2020. Comparative plastid genomics of neotropical *Bulbophyllum* (Orchidaceae; Epidendroideae) [J]. Front Plant Sci, 11: 799.
ZHANG D, GAO FL, JAKOVLIC I, et al., 2020a. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies [J]. Mol Ecol Resour, 20(1): 348 - 355.
ZHANG RS, YANG J, HU HL, et al., 2020b. A high level of chloroplast genome sequence variability in the Sawtooth Oak *Quercus acutissima* [J]. Int J Biol Macromol, 152: 340 - 348.
ZHANG YH, JIA HX, WANG ZB, et al., 2019. Genetic diversity and population structure of *Populus yunnanensis* [J]. Biodivers Sci, 27(4): 355 - 365.
ZHANG YJ, LI DZ, 2011. Advances in phylogenomics based on complete chloroplast genomes [J]. Plant Divers Resour, 33(4): 365 - 375. [张韵洁，李德铢，2011. 叶绿体系统发育基因组学的研究进展 [J]. 植物分类与资源学报，33(4):365 - 375.]

附录 1 飘带兜兰叶绿体基因组编码的基因

Appendix 1 Gene contents in the chloroplast genomes of six *Paphiopedilum parishii* individuals

基因分组 Group of genes	基因名称 Gene names
1 光合系统 I 基因 Photosystem I	*psaA, psaB, psaC* (×2), *psaI, psaJ*
2 光合系统 II 基因 Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
3 细胞色素复合物基因 Cytochrome b/f complex	*petA, petB*, petD*, petG, petL, petN*
4 ATP 合酶基因 ATP synthase	*atpA, atpB, atpE, atpF*, atpH, atpI*
5 NADH 脱氢酶基因 NADH dehydrogenase	*ndhB* * (×2), *ndhD*^Δ^, *ndhJ*^Δ^
6 二磷酸核酮糖羧化酶大亚基基因 RubisCO large subunit	*rbcL*
7 RNA 聚合酶亚基基因 RNA polymerase	*rpoA, rpoB, rpoC*1*, *rpoC*2
8 核糖体小亚基基因 Small subunit ribosome (SSU)	*rps*2, *rps*3, *rps*4, *rps*7(×2), *rps*8, *rps*11, *rps*12(×2), *rps*14, *rps*15(×2), *rps*16*, *rps*18, *rps*19(×2)
9 核糖体大亚基基因 Large subunit of ribosome (LSU)	*rpl*2*(×2), *rpl*14, *rpl*16*, *rpl*20, *rpl*22, *rpl*23(×2), *rpl*32, *rpl*33, *rpl*36
10 其他基因 Other genes	*clpP**, matK, accD*, ccsA, infA*
11 未知功能基因 Proteins of unknown function	*ycf*1(×2), *ycf*2(×2), *ycf*3**, *ycf*4, *ycf*15(×2)^Δ^
12 核糖体 RNA 基因 Ribosomal RNAs	*rrn*4.5(×2), *rrn*5(×2), *rrn*16(×2), *rrn*23(×2)
13 转运 RNA 基因 Transfer RNAs	*trnA-UGC**(×2), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnG-UCC**, *trnH-GUG* (×2), *trnI-CAU (×2)*, *trnI-GAU**(×2), *trnK-UUU**, *trnL-CAA* (×2), *trnL-UAA**, *trnL-UAG* (×2), *trnM-CAU*, *trnN-GUU* (×2), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG* (×2), *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC* (×2), *trnV-UAC**, *trnW-CCA*, *trnY-GUA*

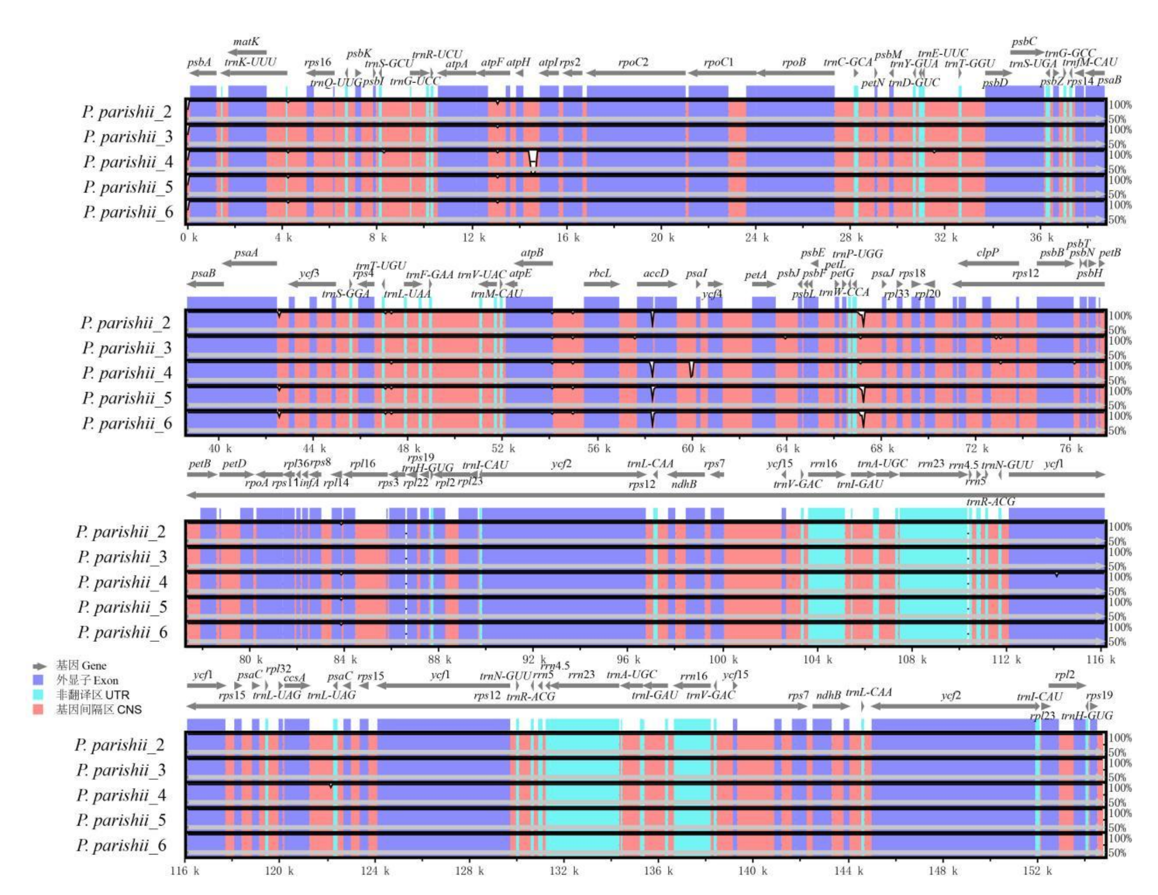
注：*表示具有一个内含子的基因；**具有两个内含子的基因；(×2)有两个拷贝的基因；Δ假基因。

Note：* indicates gene containing one intron; ** indicates gene containing two introns; (×2) indicates gene has two copies; Δ indicates pseudogene.

附录 2 飘带兜兰 6 个个体的叶绿体基因组 SSR 位点类型和数量

Appendix 2 Types and amounts of simple sequence repeats (SSRs) identified in the chloroplast genomes of six *Paphiopedilum parishii* individuals

SSR 类型 SSR Type	*P. parishii*_1		*P. parishii*_2		*P. parishii*_3		*P. parishii*_4		*P. parishii*_5		*P. parishii*_6	
	数目 Number	比例 Ration (%)	数目 Number	比例 Ration (%)	数目 Number	比例 Ration (%)	数目 Number	比例 Ration (%)	数目 Number	比例 Ration (%)	数目 Number	比例 Ration (%)
单核苷酸重复 Mononucleotide	51	48	47	46	51	48	47	46	48	46	48	46
二核苷酸重复 Dinucleotide	20	19	20	19	20	19	20	19	20	19	20	19
三核苷酸重复 Trinucleotide	15	14	15	15	15	14	15	15	15	14	15	14
四核苷酸重复 Tetranucleotide	14	13	14	14	14	13	14	14	14	13	14	13
五核苷酸重复 Pentanucleotide	2	2	2	2	2	2	2	2	2	2	2	2
六核苷酸重复 Hexanucleotide	5	5	5	5	5	5	5	5	5	5	5	5
总数 Total	107		103		107		103		104		104	

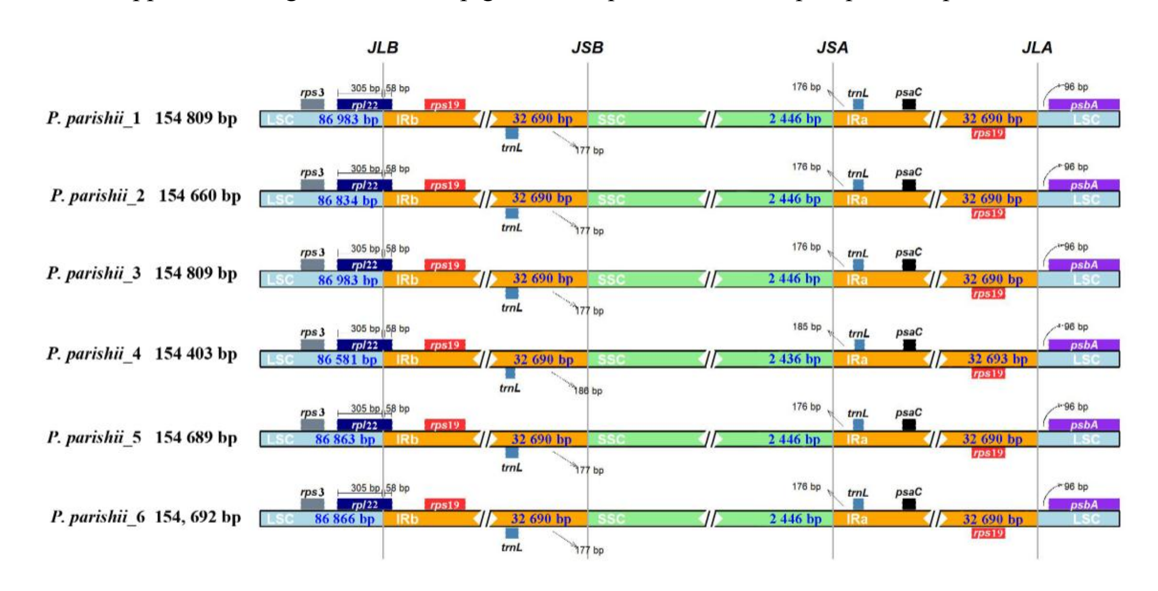


参考序列为 *Paphiopedilum parishi*_1。*X* 轴对应叶绿体基因组内的坐标。*Y* 轴显示 50%~100%范围内的百分比。

With *Paphiopedilum parishii*_1 as a reference. The *X*-axis corresponds to coordinates within the cp genome. The *Y*-axis shows the percentage identity in the range 50%-100%.

附录 3 飘带兜兰叶绿体基因序列比对

Appendix 3 Alignment of the cp genome sequences of six *Paphiopedilum parishii*

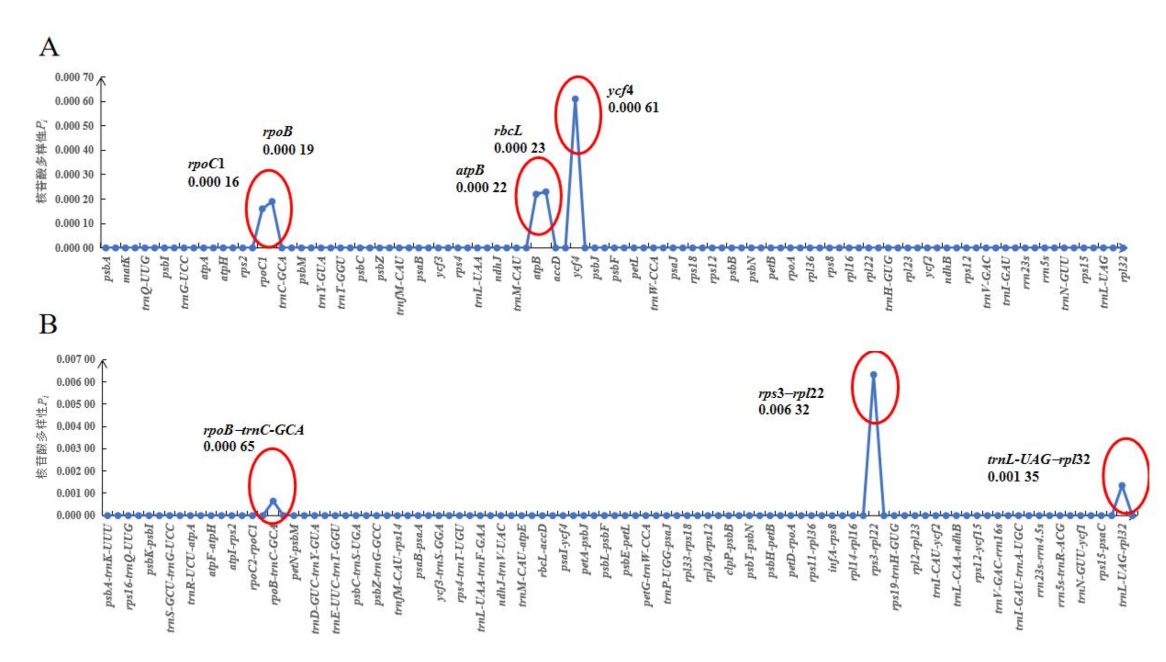


JLB, *JSB*, *JSA*, *JLA* 分别是 LSC/IRb, SSC/IRb, SSC/IRa 和 LSC/IRa 的边界。

JLB, JSB, JSA, JLA refer to junction boundary between LSC/IRb, SSC/IRb, SSC/IRa and LSC/IRa respectively.

附录 4 飘带兜兰 6 个个体叶绿体全基因组 IR 边界比较图

Appendix 4 Comparison of LSC, IRs, and SSC boundaries among the six *Paphiopedilum parishii* plastome



A. 基因编码区核苷酸多样性；**B**. 基因间隔区核苷酸多样性。

A. Nucleotide diversity (P_i) values of coding genes in the LSC, SSC, and IR regions; **B**. Nucleotide diversity (P_i) values of intergenic genes in LSC, SSC, and IR regions.

附录 5 飘带兜兰叶绿体基因组核苷酸多样性比对分析

Appendix 5 Comparative analysis of nucleotide diversity values (P_i) among the cp genome sequences of six *Paphiopedilu. parishii* individuals